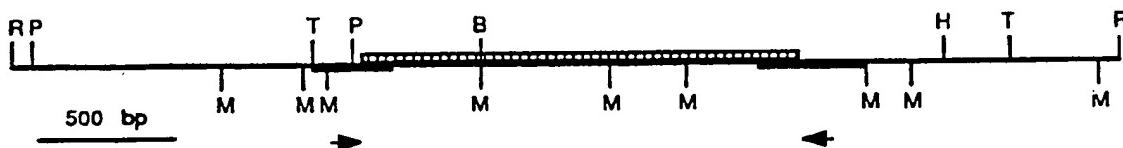


**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : <b>C12Q 1/68, C07H 21/04 C12P 19/34, C12N 15/11</b>		A2	(11) International Publication Number: <b>WO 91/14003</b> (43) International Publication Date: 19 September 1991 (19.09.91)
<b>(21) International Application Number:</b> PCT/US91/01628 <b>(22) International Filing Date:</b> 11 March 1991 (11.03.91)  <b>(30) Priority data:</b> 491,286 9 March 1990 (09.03.90) US		<b>(81) Designated States:</b> AT (European patent), BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>	
<b>(71) Applicant:</b> IG LABORATORIES, INC. [US/US]; One Mountain Road, Framingham, MA 01701 (US). <b>(72) Inventors:</b> HORN, Glenn, T. ; 169-B-11 Broadmeadow Road, Marlboro, MA 01725 (US). RICHARDS, Brenda, L. ; 91 East Main Street, Northboro, MA 01532 (US). <b>(74) Agents:</b> GRANAHAN, Patricia et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US).			

**(54) Title:** CHARACTERIZATION AND ANALYSIS OF POLYMORPHIC VNTR LOCI**(57) Abstract**

This invention pertains to characterization of the YNZ32 VNTR locus and to methods for detecting polymorphisms at a highly polymorphic VNTR locus. Polymorphisms are detected by VNTR locus amplification and enzymatic digestion of the amplification products. Variations observed in restriction maps of the fragments are then used to assess the polymorphisms. Primers for VNTR amplification of the YNZ32 locus are described.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BP	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

CHARACTERIZATION AND ANALYSIS OF  
POLYMORPHIC VNTR LOCI

Background of the Invention

The human genome is interspersed with segments of  
5 directly repeated DNA. Some of these loci vary in the  
number of tandem repeats and have been designated  
variable number of tandem repeats (VNTR). Nakamura Y.  
et al., Science 235:1616-1622 (1987). These repeated  
regions can be retrieved from human cosmid libraries by  
10 screening with oligonucleotides from known VNTRs, then  
further localizing the repeat on a smaller fragment.  
Nakamura, Y. et al. Ibid. Many of the probes for VNTR  
loci reveal more than two alleles; their extensive  
polymorphism makes them useful in genetic mapping,  
15 clinical linkage analysis, and DNA fingerprinting.  
However, few of these loci have been characterized in  
detail.

VNTR polymorphisms are currently analyzed as  
restriction fragment length polymorphisms (RFLPs) via  
20 Southern blot transfer and hybridization to radioactive  
probes. Although Southern transfer is used to detect  
many kinds of polymorphisms, it requires large amounts  
of sample DNA that is able to be digested by restriction  
enzymes, and can take up to a week to complete.

25 A VNTR segment designated pYNZ32 (human gene  
mapping (HGM) locus D4S125) was isolated as part of a  
family of cosmid clones hybridizing to a synthetic  
oligonucleotide probe from the zeta globin repeat  
sequence. Nakamura et al., Science 235:1616-1622  
30 (1987). The pYNZ32 probe was reported to reveal a  
six-allele VNTR identified by several enzymes,

- 2 -

including TaqI and PstI. Nakamura Y. et al., Nucl. Acids Res. 16(9):4186 (1988). It was described as having a heterozygosity of 58% with TaqI in 94 unrelated Caucasians and was later closely linked to 5 the locus responsible for Huntington's disease, located on the short arm of chromosome 4. Nakamura Y. et al., Science 235:1616-1622 (1987); and MacDonald, M.E. et al., J. Clin. Invest. 84:1013-1016 (1989). As such, 10 the pYNZ32 probe has been used in restriction fragment length polymorphism (RFLP) analysis as a linkage marker for Huntington's disease.

It would be desirable to provide an efficient method for determining polymorphisms at a VNTR locus which can be used as a diagnostic.

15 Summary of the Invention

The present invention is a method for determining polymorphisms which occur in VNTR loci, such as the YNZ32 and YNZ22 locus. Primers for YNZ32 are described. According to the method, sequences flanking 20 a VNTR locus of interest are determined and used to construct amplification primers which can direct amplification across the VNTR locus. The primers are extended using polymerase chain reaction (PCR) to produce amplification products which can be directly 25 analyzed. Variations in the length of the amplified products are then used to identify sequence polymorphisms. Sequence variations which do not affect the size of the amplification products can be further analyzed by digesting the amplification products with 30 various restriction enzymes to generate restriction fragments.

- 3 -

The invention can be used in genetic linkage studies, paternity testing, forensic identification or as a diagnostic for polymorphisms.

Brief Description of the Drawings

5       Figure 1 shows a restriction map of the pYNZ32 insert. Restriction sites are indicated as R, Rsal; P, PstI; T, TaqI; H, HindIII; B, BglII; M, MspI. The hatched portion indicates the repeat region. The double-thick lines represent the areas from which the  
10 sequence data in Figure 2 was taken. Arrows beneath the map show approximate positions for the amplification primers.

Figure 2 shows the partial nucleotide sequence of plasmid pYNZ32.

15       Figure 3 shows sequence diversity within the repeated region of pYNZ32 which is shown in Figure 2.

Figure 4A shows the results of agarose gel electrophoresis of PCR products amplified from total genomic DNA. Lanes 1-3: amplification of DNA from a father (showing allelic products designated A and F), a child (products F and G), and a mother (C and G); lanes 20 4,5,7,8: amplification of DNA from 4 unrelated individuals giving amplified products B and E, A and E, D and I, and D and G, respectively; lane 6: Rsal DNA marker fragments from bacteriophage  $\lambda$ . Sizes in bp are indicated to the right.

Figure 4B shows a Southern blot analysis of TaqI digested DNA of some of the same samples amplified above, and hybridized with probe pYNZ32. Overloading 30 of the first two lanes has slightly reduced the apparent size of the bands. The exposure of th se

-4-

lanes has been reduced photographically. Sizes in Kbp are shown on the right.

Figure 4C is a schematic showing the relative positions of the YNZ32 amplified products versus the 5  $\lambda$ DNA/RsaI marker fragments. The largest marker fragment represents a doublet of 2196 and 2183 bp, and the smallest fragment shown is 1436 bp.

Figure 5 shows MspI restriction maps for twelve YNZ32 amplified allele products. The products are 10 designated A to I (largest to smallest), and some products of the same size are shown which differ in their MspI digestion pattern. The alleles are aligned at the conserved MspI site, the second site in most of the alleles (except for I which has only one MspI 15 site). A star indicates the presence of the unique BglII site.

#### Detailed Description of the Invention

This invention pertains to methods for amplifying polymorphic VNTR loci and to methods for determining 20 polymorphisms which occur at the VNTR locus. By polymerase chain reaction (PCR) techniques, a segment of the VNTR locus can be amplified and its amplification products analyzed. As a result of this method, it is possible to identify sequence polymorphisms by 25 determining size variations in the PCR products.

According to the method, nucleotide sequences flanking a VNTR locus of interest are first determined. From that information, two oligonucleotide primers are designed to direct amplification across the VNTR 30 region. One oligonucleotide primer comprises a nucleotide sequence which hybridizes to the region

-5-

flanking one end of the VNTR sequence. The second primer comprises a nucleotide sequence which hybridizes to the region of flanking the other end of the VNTR sequence. Hybridization of the primers to other 5 portions of the VNTR can be prevented by restricting the primers to those nucleotides which are unique to the flanking sequences and avoiding segments of DNA which are partially homologous to the repeat unit.

The primers are then subjected to conditions 10 sufficient to direct amplification across the VNTR locus. This can be accomplished using PCR techniques described by Mullis, K.B. et al., U.S. Patent No. 4,683,195 and Mullis, K.B., U.S. Patent No. 4,683,202.

15 The teachings of these patents are incorporated by reference herein. The resulting amplification products can then be analyzed for size variations using gel electrophoresis techniques. Variations observed in the length of the fragments are indicative of sequence polymorphisms.

20 Sequence polymorphisms which do not affect the length of the PCR product can be determined by digesting them with various restriction enzymes to produce fragments from which a restriction map can be determined. The lengths of the fragments are 25 indicative of sequence polymorphisms.

The methods of this invention provide rapid analysis of polymorphisms at VNTR loci using small amounts of sample DNA. Such analysis of these regions is extremely useful in genetic linkage studies, 30 paternity testing and forensic identification. One advantage is that cumbersome RFLP-based assays are not needed.

The present invention is based on the characterization of the highly polymorphic VNTR locus pYNZ32, which had been previously isolated. Nakamura, Y. et al., Science 235:1616-1622 (1987). For the first time, 5 a restriction map for this locus has been determined, as well as the nucleotide sequence of a substantial portion of this locus. Based on this information, oligonucleotides derived from this sequence can be synthesized. Such synthetic oligonucleotides are 10 useful as primers for amplification techniques and can be used to direct amplification across the VNTR region. DNA sequencing in the areas within and flanking the repeated segment allowed for the design of specific amplification primers. The repeated region of pYNZ32 15 consists of an imperfectly duplicated 27 base pair motif, 16 bases of which are more highly conserved. Allelic products from PCR amplification were resolved into nine different size classes ranging from approximately 1400 to 2200 base pairs. Additional polymorphism was revealed when the amplified products were 20 analyzed by restriction enzyme digestion. Both the overall size variation as well as the internal sequence polymorphism were used to determine a heterozygosity value of 86% for YNZ32 in 50 unrelated individuals. 25 The rapid analysis and improved resolution of amplified alleles on agarose gels, and the internal variability within YNZ32, increase its diagnostic utility as a VNTR and as a linkage marker for the nearby Huntington's disease gene.

- 7 -

Characterization of the YNZ32 Locus

The cloned insert in probe pYNZ32 was mapped for various restriction sites (Figure 1). The distance between the two TaqI sites suggests they are 5 responsible for the known TaqI RFLP (2.3-2.8 Kbp) for YNZ32 on human genomic DNA Southern blots. Nakamura Y. et al., Science 235:1616-1622 (1987). Similarly, the PstI site is most likely involved with the PstI polymorphism, with the other site lying further to the 10 right of the segment contained within the cloned fragment. In contrast to a previously published MspI RFLP range of 1.2-1.8 Kbp for YNZ32 on genomic DNA Southern blots (Nakamura Y. et al., Nucl. Acids Res. 16:4186 (1988)), four fragments in the 400-700 bp range 15 were observed, consistent with the map of the cloned probe (Figure 1).

Sequencing of MspI fragments within the clone revealed the tandemly repeated segment, which is approximately 1700 base pairs long and totally contained within the cloned insert (Figure 1). The DNA sequences near the termini of the repeated region are shown in Figure 2, and additional sequencing was performed on internal fragments (Figure 3). This region shows strong purine/pyrimidine asymmetry, and 25 predominantly consists of tetranucleotides CCTT and CATT. Consideration of the arrangement of these tetranucleotides suggests a 27 bp repeated motif (Figure 3), 16 bp of which are more highly conserved. However, no part of this region shows the exact tandemly repeated segments seen in other VNTR loci. 30 Jeffreys, A.J. et al., Nature 332:278-281 (1988); and Wolff, R.K. et al., Genomics 3:347-351 (1988).

- 8 -

#### Amplification and Analysis

The sequences flanking the repeat region (Figure 2) were also characterized. The information was used to construct primers for amplification of the repeat 5 region. Primer GH301 and GH290 were used to amplify the YNZ32 locus from human genomic DNA samples using polymerase chain reaction techniques. The amplification products ranged in size from approximately 1400 to 2200 base pairs. These products were grouped into 9 10 size classes designated A to I in order of decreasing size (amplification from the pYNZ32 clone yielded a "D" allele), according to their positions relative to the fragments in RsaI digested lambda DNA (Figures 4A and 4C). Each size class encompasses a narrow range of 15 amplified allelic products differing slightly in length. Sizes of the PCR products from specific samples correlate with the results seen in Southern blot analysis (Figure 4B), thus validating the use of PCR for detecting polymorphisms at the YNZ32 locus.

20 The Mendelian inheritance of the amplified alleles is also verified in Figure 4B.

The minor variations seen within these allelic size classes indicate that additional polymorphism might be elucidated by restriction enzyme mapping.

25 Several amplified products were therefore isolated, reamplified, and analyzed by digestion with MspI or BglII. Twelve of these products having different digestion patterns were also mapped with MspI (Figure 5). Most of the products have three MspI sites; the 30 maps in Figure 5 are lined up at the second MspI site, which is conserved in all but one of the alleles. Allelic products A1 and F1 are missing the first MspI

- 9 -

site, while products B1, C1 and G1 are missing the third MspI site. The I1 allele has only one internal MspI site. The unique BglI site shown on the map in Figure 1 is present in only some alleles. Finally, the 5 MspI fragments show length variations even within an allelic size class. For example, the right-hand fragment produced by MspI digestion is 555 bp in product D1 but only 500 bp in D2 (Figure 5).

The frequencies of alleles A to I in 50 unrelated 10 individuals were determined (Table 1). The largest allele (A) and the two smallest alleles (H and I) were each present in only one of these 100 chromosomes. Almost 80% of the products were found to be in the intermediate classes of D, E and F. The distribution 15 of alleles among these individuals was consistent with Hardy-Weinberg equilibrium, and no evidence for null alleles was observed. YNZ32 was found to have a heterozygosity value of 86% for these 100 chromosomes.

-11-

or the zeta-globin oligonucleotide with which both of these probes were selected. Nakamura Y. et al., Science 235:1616-1622 (1987). A search of sequences listed in GenBank revealed the greatest similarity to a 5 polypyrimidine repeat isolated from the mouse genome. Deugau, K.V. et al., Anal. Biochem. 129:88-97 (1983).

Amplification of the YNZ32 locus by PCR is a rapid alternative to analysis by other assays. Both the amplification and agarose gel analysis of YNZ32 can be 10 performed within four hours. The direct visualization of PCR products on ethidium bromide-stained agarose gels allows the allelic products to be better resolved than the bands seen on Southern blots. Amplified products are also amenable to restriction enzyme 15 analysis, which reveals additional polymorphism.

In addition to the overall variation in length of the repeated region, the YNZ32 locus contains polymorphic restriction sites. This additional polymorphism can be used to discriminate alleles of similar 20 size, thereby increasing the diagnostic utility of this locus. For example, the B1 and B2 alleles (Figure 5) give the same size PCR products, but when digested with MspI, release distinctive fragments of 775 or 605 bp, respectively. Similar discrimination between certain 25 alleles can also be achieved using the single BglII site, which is present only in some chromosomes.

The ability to rapidly amplify and analyze this highly polymorphic locus should help support both clinical and research studies. The extensive heterozygosity of YNZ32 alleles is often critical in 30 discriminating chromosomes in the large pedigrees used in linkage analysis of Huntington's disease.

-12-

MacDonald, M.E. et al., J. Clin. Invest. 84:1013-1016 (1989). This polymorphism may also be of use in more general 'DNA fingerprinting' studies. Finally, the divergent nature of the repeated sequence at this locus 5 may assist in understanding the creation, evolution, and stability of VNTR segments. Jarman, A.P. and R.A. Wells, Trends in Genet. 5:367-371 (1989).

The invention will be further illustrated by the following exemplification, which is not intended to be 10 limiting in any way.

#### EXEMPLIFICATION

##### Materials and Methods

The clone pYNZ32 was obtained from Y. Nakamura and is also available from the ATCC (#57548). It has a 15 total insert size of 4.1 Kbp. High molecular weight genomic DNA used for amplification was prepared by standard methods (Maniatis T. et al., "Molecular Cloning - A Laboratory Manual", Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1982)).

20 Oligonucleotide primers were synthesized on an Applied Biosystems 380B Synthesizer. Agarose gels were made from SEAKEM GTG agarose (American Bioanalytical) and were run in 1X TBE buffer with 0.1 µg/ml ethidium bromide at 10 volts per cm.

##### 25 Mapping and Sequencing

The pYNZ32 clone was mapped with several restriction enzymes. The positions of MspI sites were more accurately determined by partial MspI digestion of the clone after labeling at one end with  $\alpha$ -<sup>32</sup>P-dATP.

-13-

Smith, H.O. and M.L. Bernstiel, Nucl. Acids Res. 3:2387 (1976). MspI fragments of pYNZ32 were subcloned into M13mpl8 and sequenced using standard dideoxy methods.

5 Sambrook, J. et al., "Molecular Cloning - A Laboratory Manual", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). Figure 3 shows the sequence diversity within the repeated region of pYNZ32. The upper line shows the sequence of the proposed 27 base pair repeated motif, with the more

10 conserved 16 base pair region underlined. Below it are segments from the pYNZ32 sequence, aligned to show similarities with the repeated core. The exact positions of some of the aligned segments within pYNZ32 are not known, but are generally listed from left to

15 right relative to Figure 1. A dash indicates identity with the core, and spaces have been inserted for two segments to improve the alignment. An asterisk indicates that this segment is immediately followed by another segment similar to the motif. Figure 2 shows

20 the results of sequencing the areas flanking the repeated region. PCR primers GH290 and GH301 were designed from these regions and have the following nucleotide sequences:

GH290: 5'-GTCAGCACCTCAGACCAAG-3'  
25 GH301: 5'-CTGTGGGACCTCGGCTAAC-3'

DNA fragments flanking and within the repeated region of pYNZ32 (Figure 1) were subcloned and sequenced (Figure 2). The first segment shown here runs from the left TaqI site until about 130 base pairs  
30 into the repeat. Also shown are the locations of the

-14-

MspI and PstI restriction sites used in RFLP analysis. The arrow indicates the location and direction of the GH301 oligonucleotide primer used in PCR amplification. The repeated region begins just after the PstI site, 5 and the more highly conserved core segment (Figure 3) is underlined. The second segment shows the final 130 base pairs of the repeat, the location and orientation of the GH290 primer, and the sequence flanking the repeated region. Additional sequence was determined 10 between these segments and is include in Figure 3, but could not be accurately placed due to the repetitive nature of this region.

#### Amplification and Analysis

PCR amplification reactions were performed using 15 AmpliTaq DNA polymerase and a DNA Thermal Cycler (Perkin Elmer). One  $\mu$ g of human genomic DNA was amplified for 28 cycles of 10 seconds at 94°C, 10 seconds at 55°C, and 30 seconds at 74°C, with a final soak for 10 minutes at 74°C. The reaction buffer used 20 was described previously (Saiki, R.K. et al., Science 239:487-491 (1988)), with the exception that KCl was not included, and the concentration of MgCl<sub>2</sub> was reduced to about 1.2mM. The amplification should be optimized empirically. The yield of PCR product was 25 also reduced when amplification primers were used at less than 1 $\mu$ M.

PCR products were analyzed on 1% agarose gels. A RsaI digest of bacteriophage  $\lambda$  DNA was used as markers for the amplified alleles since it has several 30 fragments in the appropriate size range. Amplified allelic products are designated A to I (largest to

-15-

smallest) by their positions relative to these size markers (Figure 4C). Those products within a single size class but having internal variation are distinguished by numbers (e.g., product B1 vs. B2).

5 Further analysis of the amplified DNA was performed by electroeluting the DNA from a gel slice (Maniatis, T. *et al.*, "Molecular Cloning - A Laboratory Manual", Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1982)), concentrating a DNA on the  
10 Centricon-30 microconcentrator (Amicon), and reamplifying individual products and analyzing their MspI and BgII digestion patterns on 2% and 1% agarose gels, respectively. In addition, 12 individually amplified products were also mapped (Figure 5) using  
15 MspI partial digestion. Smith, H.O. and M.L. Bernstiel, Nucl. Acids Res. 3:2387 (1976).

Five DNA samples that were amplified by PCR were also analyzed as TaqI digests on a Southern blot.

Maniatis, T. *et al.*, "Molecular Cloning - A Laboratory  
20 Manual", Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1982). The blot (Figure 4B) was hybridized with the 2900 bp PstI-RsaI fragment of pYNZ32 shown on the map of Figure 1.

#### Equivalents

25 Those skilled in the art will recognize or be able to ascertain, using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. These and all other equivalents are intended to be encompassed by the  
30 following claims.

-16-

CLAIMS

1. A method of detecting polymorphisms at a polymorphic variable number of tandem repeats (VNTR) locus, comprising the steps of:
  - a) providing two oligonucleotide primers which specifically direct amplification across the VNTR locus, wherein one of the primers comprises a nucleotide sequence which hybridizes to the region flanking one end of the VNTR sequence and the other of the primers comprises a nucleotide sequence which hybridizes to the region flanking the other end of the VNTR sequence;
  - b) contacting the two primers with genomic DNA thought to contain the VNTR of interest, under conditions sufficient to direct amplification across the VNTR locus and produce amplification products thereof; and
  - c) determining size variations in the amplification products, wherein the variations are indicative of sequence polymorphisms.
2. The method of Claim 1, wherein the amplification is performed by polymerase chain reaction.
3. The method of Claim 2, wherein the VNTR locus is YNZ32 and the oligonucleotide primers have nucleotide sequences shown in Figure 2 or portion thereof, for directing amplification across the YNZ32 locus.

-17-

4. The method of Claim 3, the oligonucleotide primers have the sequences 5'-GTCAGGACCTCAGACCAGAG-3' and 5'-CTGTGGGACCTCGGCTAACCC-3'.

5. The method of Claim 1, further comprising the step  
of:  
d) digesting the amplification products with restriction enzymes to produce DNA fragments having lengths which correspond to sequence polymorphisms.

10 6. A method of amplifying a polymorphic VNTR locus, comprising the steps of:  
a) providing two oligonucleotide primers which specifically direct amplification across the VNTR locus, wherein one of the primers  
15 comprises a nucleotide sequence which hybridizes to the region flanking one end of the VNTR sequence and the other of the primers comprises a nucleotide sequence which hybridizes to the region flanking the other end of  
20 the VNTR sequenced; and  
b) hybridizing the primers to genomic DNA thought to contain the VNTR of interest, under conditions sufficient to direct amplification across the VNTR locus and produce  
25 amplification products thereof.

7. The method of Claim 6, wherein the amplification is performed by polymerase chain reaction.

8. The method of Claim 7, wherein the VNTR locus is YNZ32 and the oligonucleotide primers have

-18-

nucleotide sequences shown in Figure 2 or portion thereof, for directing amplification across the YNZ32 locus.

9. The method of Claim 8, the oligonucleotide primers  
5 have the sequences 5'-GTCAGCACCTCAGACCAGAG-3' and  
5'-CTGTGGGACCTCGGCTAAGC-3'.
10. An oligonucleotide primer having the nucleotide sequences shown in Figure 2 or portion thereof, for directing amplification across the YNZ32 locus.
11. The primer of Claim 10, having the sequences  
5'-GTCAGCACCTCAGACCAGAG-3' or  
5'-CTGTGGGACCTCGGCTAAGC-3'.

- 1 / 6 -

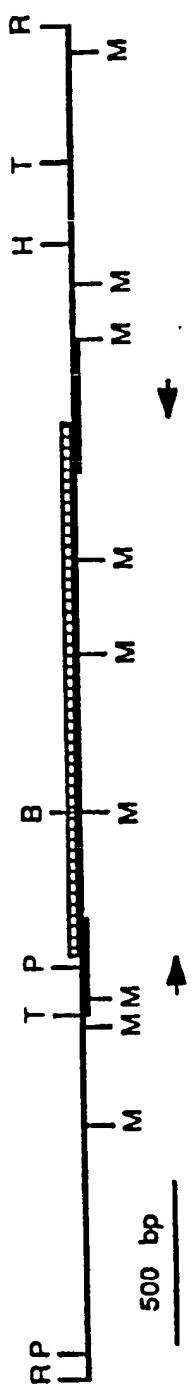


FIG. 1

- 2 / 6 -

(TaqI)

GACCCCTGACTGTGAGGGCGTGTGCCAGGAAGCCCTGGAACCCCTACAGCTGGAGA  
CTGGGAGCTGACACTCCTCGCACACACGGTCTGGACCTTTGGATGTCGACCTCT

(MsPI)

ACGCTGAGCCTCCTCACATGGGGCAGGGGGCACTGGAGGTAGGGGGGGCTGACC  
TGGGACTCGGAGGAGTGTACGGCCGTACGGCTTACGGCTTACCTCCATCCCCGGACTGG

GH301 →

TGGGCCTGGGACCTGGCTTAAGCCTCCAGCAGGAGCTCAGGAGGCATCCATGCATGT  
ACCGAACCCCTGGAGGCCATTGGAGGTCTGGAGGTCTCGAGTCCTCCGTAGGTACGTACA

(Pst I)

GTGGGGCTGGCTCCCTCTGCTCATTCATTCATCTCATTCATTCATCTCCTGCTCATTCATT  
CACCCCGACGTCGGAGGGAGACGAGTAAGTAAGGAAGAACAGGGACGAGGACGAGTAAGTAA

CTTCCCNCAACTCCCCGGCTACTCATTCATCCCTCCTCCCCACCACCCG  
GAAGGGNGTTGAGGGGGATGAGTAAGTAAGTAAGGAAGAACAGGGACGAGGACGAGTAAGTAA

CTCATTCCTTGCCTCTGGCCA ..... (approx. 1600 bp) .... GTTCC  
GAGTAAGGAACGGAAAGAGAGACGGGT CAAGG

CCTGGCTCATTCATCTCTCTNNCTACTCAATTCTCTCACTCATTCATCTCTCT  
GGACGAGTAAGTAAGGAAGAACAGANNGATGAGTTAAAGAGAGTGAGTAAAGTAAGA

FIG.2 (SHEET 1 OF 2)

- 3 / 6 -

TTCTTGTCCCTCTGCTTATTCTCCTTCCTGTTCTGCATTCAATTCTTCCTTC  
AAGAACAGGGAGACGAATAAGGAAGGAAGGAACAAGTAAGTAAGGAAGT  
  
CTTATGGTTCACATTACTAAACTCTGGTCTGAGGGCTTGACGGAAAGGTGGTTCTGC  
GAATAACCAAGGTGTTAAATGATTGAGCCAGACTCCACGACTGCCCCTTCCACAAAGACG  
  GH290 →  
  
CCTGAGAGGCCCTTGAATCCCTTCCCTGCCACACCTTGCTGGGGAGACCAAGGCC  
GGACTCTCGAAACTTAGGAAGGACGGGTCTGTCAACACCGACCCCTCTGGTTCCGGGA  
  
GCATGGTGGTTTACCTTCTGTGCCAGACAAAGGCTCTAAGGCAGGTGAGCTGGCTGC  
CGTACAAACCAAATGAAAGACACGGGTCTGTGGAGATTCCGTCCACTCGACCGACG  
  
(MsPI)  
CCAGGAGGAGGGCACCTCCTCAAGCTGGCC  
GGTCCTCGTCTCCGGTGGAGTCGACGGCC

FIG. 2 (SHEET 2 OF 2)

- 4 / 6 -

### Sequence Diversity within pYNZ32

TGCTCATTCATTCCCTTCTCCCC  
-----A-----GT-T-C \*  
-----T-C--NCAA-T---  
-A----- C-A--  
CCT-----T-T--- \*  
-----C-A-T \*  
-----T-----CNG--- \*  
CCTC-----T-----A \*  
CA-----TC-----A  
-----G-----A-TC--- \*  
A-----C-----T-T- \*  
-----C-T-----  
G-----C-----TCC-CA-T-- \*  
---A-----A-----C----- \*  
-----T-C-C-GCTAT-T  
-----T-----A-----T-- \*  
-----A-----ATT-TT  
-----T-G--  
---TG---C-----CTGTT---C \*  
-----G----- \*  
-----GGCT-A--  
-----T-----TNN-  
CA-----T-T-G-----T- \*  
---T-----C-----CTGT- \*  
---T-----T-----A-TTA

FIG. 3

- 5 / 6 -



Figure 4a



Figure 4b

Figure 4c

- 6 / 6 -

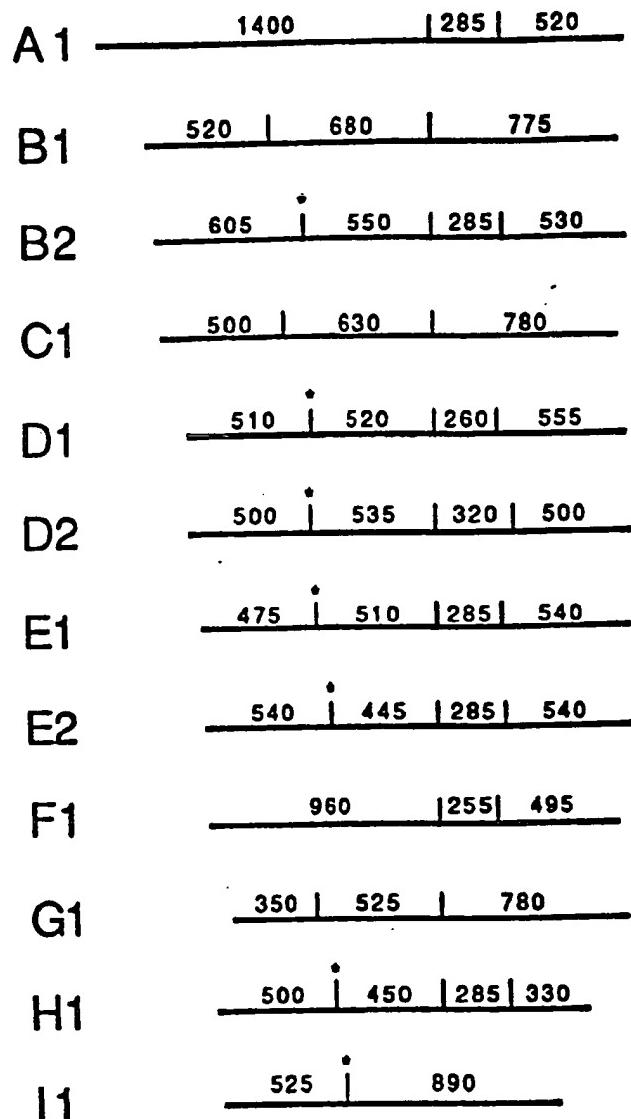


FIG. 5